

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: POTTER, ANDREW A.
REDMOND, MARK J.
HUGHES, HUW P.A.
- (ii) TITLE OF INVENTION: ENHANCED IMMUNOGENICITY USING LEUKOTOXIN CHIMERAS
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: ROBERTA L. ROBINS
(B) STREET: 635 BRYANT STREET
(C) CITY: PALO ALTO
(D) STATE: CALIFORNIA
(E) COUNTRY: UNITED STATES OF AMERICA
(F) ZIP: 94301
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patent In Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/960,932
(B) FILING DATE: 14-OCT-1992
(C) CLASSIFICATION: 435
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: ROBINS, ROBERTA L.
(B) REGISTRATION NUMBER: 33,208
(C) REFERENCE/DOCKET NUMBER: 9000-0016.20
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 617-8999
(B) TELEFAX: (415) 327-3231

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2794 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..2778

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGCTACTG TTATAGATCT AAGCTTCCCA AAAACTGGGG CAAAAAAT TATCCTCTAT	60
ATTCCCCAAA ATTACCAATA TGATACTGAA CAAGSTAATG GTTTACAGGA TTTAGTCAAA	120
GCGGCCGAAG AGTTGGGGAT TGAGGTACAA AGAGAAGAAC GCAATAATAT TGCAACAGCT	180

CAAACCAAGTT TAGGCACGAT TCAAACCGCT ATTGGCTTAA CTGAGCGTGG CATTGTGTTA	240
TCCGCTCCAC AAATTGATAA ATTGCTACAG AAAACTAAAG CAGGCCAAGC ATTAGGTTCT	300
GCCGAAAGCA TTGTACAAA TGCAATAAA GCCAAACTG TATTATCTGG CATTCAATCT	360
ATTTTAGGCT CAGTATTGGC TGGGAATGGAT TTAGATGAGG CCTTACAGAA TAACAGCAAC	420
CAACATGCTC TTGCTAAAGC TGGCTTGGAG CTAACAAATT CATTAAATTGA AAATATTGCT	480
AATTCAGTAA AAACACTTGA CGAATTTGGT GAGCAAATTA GTCAATTTGG TTCAAACTA	540
CAAAATATCA AAGGCTTAGG GACTTTAGGA GACAACTCA AAAATATCGG TGGACTTGAT	600
AAAGCTGGCC TTGGTTTAGA TGTTATCTCA GGGCTATTAT CGGGCGCAAC AGCTGCACTT	660
GTAATTGCTC ATAAAAATGC TTCAACAGCT AAAAAAGTGG GTGCGGGTTT TGAATTGGCA	720
AACCAAGTTG TTGGTAATAT TACCAAGCC GTTCTTCTT ACATTTTAGC CCAACGTGTT	780
GCAGCAGSTT TATCTTCAAC TGGGCCTGTG GCTGCTTTAA TTGCTTCTAC TGTTCCTCTT	840
GCGATTAGCC CATTAGCATT TGCCGGTATT GCCGATAAAT TTAATCATGC AAAAGSTTA	900
GAGAGTTATG CCGAACGCTT TAAAAATTA GGCTATGACG GAGATAATTT ATTACAGAA	960
TATCAGCGGG GAACAGGGAC TATTGATGCA TCGGTTACTG CAATTAATAC CGCATTGGCC	1020
GCTATTGCTG GTGGTGTGTC TGCTGCTGCA GCCGGCTCGG TTATTGCTTC ACCGATTGCC	1080
TTATTAGTAT CTGGGATTAC CGGTGTAATT TCTACGATTC TGCAATATTC TAAACAAGCA	1140
ATGTTTGAGC ACGTTGCAA TAAATTCAT AACAAATTG TAGAATGGGA AAAAAATAAT	1200
CACGGTAAGA ACTACTTTGA AAATGGTTAC GATGCCCGTT ATCTTGCGAA TTTACAAGAT	1260
AATATGAAAT TCTTACTGAA CTAAACAAA GAGTTACAGG CAGAACGTGT CATCGCTATT	1320
ACTCAGCAGC AATGGGATAA CAACATTGGT GATTAGCTG GTATTAGCCG TTTAGGTGAA	1380
AAAGTCCTTA GTGGTAAAGC CTATGTGGAT GCGTTTGAAG AAGGCAACA CATTAAAGCC	1440
GATAAATTAG TACAGTTGGA TTCGGCAAAC GGTATTATTG ATGTGASTAA TTCGGGTAAA	1500
GCGAAACTC AGCATATCTT ATTCAGAACG CCATTATTGA CGCCGGGAAC AGAGCATCGT	1560
GAACGCGTAC AAACAGGTAA ATATGAATAT ATTACCAAGC TCAATATTAA CCGTGTAGAT	1620
AGCTGGAAAA TTACAGATGG TGCAGCAAGT TCTACCTTTG ATTTAACTAA CGTTGTTTAC	1680
CGTATTGGTA TTGAATTAGA CAATGCTGGA AATGTAATA AAACCAAAGA AACAAAAATT	1740
ATTGCCAAAC TTGGTGAAGG TGATGACAAC GTATTTGTTG GTTCTGGTAC GACGGAATT	1800
GATGGCGGTG AAGGTTACGA CCGAGTTCAC TATAGCCGTG GAACTATGG TGCTTTAACT	1860
ATTGATGCAA CCAAAGAGAC CGAGCAAGST AGTTATACCG TAAATCGTTT CGTAGAAACC	1920
GGTAAAGCAC TACACGAAGT GACTTCAACC CATACCGCAT TAGTGGGCAA CCGTGAAGAA	1980
AAAATAGAAT ATCGTCATAG CAATAACCAG CACCATGCCG GTTATTACAC CAAAGATACC	2040

TTGAAAGCTG TTGAGAAAT TATCGGTACA TCACATAACG ATATCTTTAA AGGTASTAAG 2100
 TTCAATGATG CCTTTAACGG TGGTGATGGT GTCGATACTA TTGACGGTAA CGACGGCAAT 2150
 GACCGCTTAT TTGGTGGTAA AGGCGATGAT ATTCTCGATG GTGSAATGG TGATGATTTT 2220
 ATCGATGGCG GTAAAGGCAA CGACCTATTA CACGGTGGCA AGGCGGATGA TATTTTCGTT 2280
 CACCGTAAAG GCGATGGTAA TGATATTATT ACCGATTCTG ACGGCAATGA TAAATTATCA 2340
 TTCTCTGATT CGAACTTAAA AGATTTAACA TTTGAAAAAG TTAAACATAA TCTTGTCATC 2400
 ACGAATAGCA AAAAGAGAA AGTGACCATT CAAACTGGT TCCGAGAGGC TGATTTTGCT 2460
 AAAGAAGTGC CTAATTATAA AGCAACTAA GATGAGAAAA TCGAAGAAAT CATCGGTCAA 2520
 AATGGCGAGC GGATCACCTC AAAGCAAGTT GATGATCTTA TCGCAAAAGG TAACGGCAAA 2580
 ATTACCCAAG ATGAGCTATC AAAAGTTGTT GATAACTATG AATTGCTCAA ACATACCAA 2640
 AATGTGACAA ACAGCTTAGA TAAGTTAATC TCATCTGTAA GTGCATTTAC CTCGTCTAAT 2700
 GATTCGAGAA ATGTATTAGT GGCTCCAAC TCAATGTTGG ATCAAAGTTT ATCTTCTCTT 2760
 CAATTTGCTA GGGGATCCTA GCTAGCTAGC CATG 2794

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCCAGCTC TTCTGCCGGC TCGAAAACT TCTTCTGGAA AACCTTCACC AGCTGCTAGG 60

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GATCCCTAGC AGCTGGTGAA GGTTTTCCAG AAGAAGTTTT TGCAGCCGGC AGAAGAGCTG 60

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATCTCAGCA TTGGAGCTAC GGCCTGCGCC CTGGCTAAG

39

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GATCCTTAGC CAGGGCGCAG GCCGTAGCTC CAATGCTGA

39

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCTTGCAA CATTGTGCCT GTGAGCATTG TGAGCCGCAA CATTGTGTAC ACCCGCGCGC

60

AACCTAACCA AGACATTGTG TAG

83

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCCTACAC AATGTCTTGG TTAAGTTGCG CGCGGTTGTA CACAATGTTG CGGCTCACAA

60

TCGTCACAGG CACAATGTTG CAA

83

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2838 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2829

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGCTACTG TTATAGATCT AAGCTTCCCA AAACTGGGG CAAAAAATAT TATCCTCTAT	60
ATTCCCCAAA ATTACCAATA TGATACTGAA CAAGGTAATG GTTACAGGA TTTAGTCAAA	120
GCGGCCGAAG AGTTGGGGAT TGAGGTACAA AGAGAAGAAC GCAATAATAT TGCAACAGCT	180
CAAACAGTT TAGGCACGAT TCAAACCGCT ATTGGCTTAA CTGAGCGTGG CATTGTGTTA	240
TCCGCTCCAC AAATTGATAA ATTGCTACAG AAACTAAAG CAGGCCAAGC ATTAGGTTCT	300
GCCGAAAGCA TTGTACAAAA TGCAATAAAA GCCAAACTG TATTATCTGG CATTCAATCT	360
ATTTTAGGCT CAGTATTGGC TGGATGGAT TTAGATGAGG CCTTACAGAA TAACAGCAAC	420
CAACATGCTC TTGCTAAAGC TGGCTTGGAG CTAACAATT CATTAATTGA AAATATTGCT	480
AATTCAGTAA AAACACTTGA CGAATTGGT GAGCAATTA GTCAATTTGG TTCAAACTA	540
CAAAATATCA AAGGCTTAGG GACTTTAGGA GACAACTCA AAAATATCGG TGGACTTGAT	600
AAAGCTGGCC TTGGTTTAGA TGTTATCTCA GGGCTATTAT CGGGCGCAAC AGCTGCACTT	660
GTACTTGCAG ATAAAAATGC TTCAACAGCT AAAAAGTGG GTGCGGGTTT TGAATTGGCA	720
AACCAAGTTG TTGTAATAT TACCAAAGCC GTTCTTCTT ACATTTTAGC CCAACGTGTT	780
GCAGCAGGTT TATCTTCAAC TGGGCCTGTG GCTGCTTTAA TTGCTTCTAC TGTCTCTCTT	840
GCGATTAGCC CATTAGCATT TGCCGGTATT GCCGATAAAT TTAATCATGC AAAAAGTTTA	900
GAGAGTTATG CCGAACGCTT TAAAAATTA GGCTATGACG GAGATAATTT ATTAGCAGAA	960
TATCAGCGGG GAACAGGGAC TATTGATGCA TCGTTACTG CAATTAATAC CGCATTGGCC	1020
GCTATTGCTG GTGGTGTGTC TGCTGCTGCA GCCGGCTCGG TTATTGCTTC ACCGATTGCC	1080
TTATTAGTAT CTGGGATTAC CGGTGTAATT TCTACGATTC TGCAATATTC TAPACAAGCA	1140
ATGTTTGAGC ACGTTGCAAA TAAATTCAT AACAAAATTG TAGAATGGGA AAAAAATAAT	1200
CACGGTAAGA ACTACTTGA AAATGTTTAC GATGCCCCGT ATCTTGCGAA TTTACAAGAT	1260
AATATGAAAT TCTTACTGAA CTTAAACAAA GAGTTACAGG CAGAACGTGT CATCGCTATT	1320
ACTCAGCAGC AATGGGATAA CAACATTGGT GATTTAGCTG GTATTAGCCG TTTAGGTGAA	1380
AAAGTCCTTA GTGGTAAAGC CTATGTGGAT GCGTTTGAAG AAGGCAAACA CATTAAAGCC	1440
GATAAATTAG TACAGTTGGA TTCGGCAAAC GGTATTATTG ATGTGAGTAA TTCGGGTAAA	1500
GCGAAACTC AGCATATCTT ATTCAGAACG CCATTATTGA CGCCGGGAAC AGAGCATCGT	1560
GAACGCGTAC AAACAGGTAA ATATGAATAT ATTACCAAGC TCAATATTAA CCGTGTAGAT	1620
AGCTGGAAAA TTACAGATGG TGCAGCAAGT TCTACCTTTG ATTTAACTAA CGTTGTTTCA	1680

CGTATTGGTA TTGAATTAGA CAATGCTGGA AATGTA ACTA AAACCAAAGA AACAAAATT	1740
ATTGCCAAAC TTGGTGAAGG TGATGACAAC GTATTTGTTG GTTCTGGTAC GACGGAAATT	1800
GATGGCGGTG AAGGTTACGA CCGAGTTCAC TATAGCCGTG GAAACTATGG TGCTTTAACT	1860
ATTGATGCAA CCAAAGAGAC CGAGCAAGGT AGTTATACCG TAAATCGTTT CGTAGAAACC	1920
GGTAAAGCAC TACACGAAGT GACTTCAACC CATACCGCAT TAGTGGGCAA CCGTGAAGAA	1980
AAAATAGAAT ATCGTCATAG CAATAACCAG CACCATGCCG GTTATTACAC CAAAGATACC	2040
TTGAAAGCTG TTGAAGAAAT TATCGGTACA TCACATAACG ATATCTTTAA AGGTAGTAAG	2100
TTCAATGATG CCTTTAACGG TGGTGATGGT GTCGATACTA TTGACGGTAA CGACGGCAAT	2160
GACCGCTTAT TTGGTGGTAA AGGCGATGAT ATTCTCGATG GTGGAAATGG TGATGATTTT	2220
ATCGATGGCG GTAAAGGCAA CGACCTATTA CACGGTGGCA AGGGCGATGA TATTTTCGTT	2280
CACCGTAAAG GCGATGGTAA TGATATTATT ACCGATTCTG ACGGCAATGA TAAATTATCA	2340
TTCTCTGATT CGAACTTAAA AGATTTAACA TTTGAAAAG TTAAACATAA TCTTGTCTATC	2400
ACGAATAGCA AAAAGAGAA ACTGACCATT CAAACTGGT TCCGAGAGGC TGATTTTGCT	2460
AAAGAAAGTC CTAATTATAA ACCAACTAAA GATGAGAAAA TCGAAGAAAT CATCGGTCAA	2520
AATGGCGAGC GGATCACCTC AAAGCAAGTT GATGATCTTA TCGCAAAAGG TAACGGCAAA	2580
ATTACCCAAG ATGAGCTATC AAAAGTTGTT GATAACTATG AATTGCTCAA ACATAGCAAA	2640
AATGTGACAA ACAGCTTAGA TAAGTTAATC TCATCTGTAA GTGCATTTAC CTCGTCTAAT	2700
GATTCGAGAA ATGTATTAGT GGCTCCAACT TCAATGTTGG ATCAAAGTTT ATCTTCTCTT	2760
CAATTTGCTA GGGGATCCAG CTCTTCTGCC GGCTGCAAAA ACTTCTTCTG GAAAACCTTC	2820
ACCAGCTGCT AGGGATCC	2838

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2817 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2808

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGCTACTG TTATAGATCT AAGCTTCCCA AAAACTGGGG CAAAAAAAT TATCCTCTAT	60
ATTCCCCAAA ATTACCAATA TGATACTGAA CAAGGTAATG GTTTACAGGA TTTAGTCAAA	120
GCGGCCGAAG AGTTGGGGAT TGAGGTACAA AGAGAAGAAC GCAATAATAT TGCAACAGCT	180

CAAACCAAGTT TAGGCACGAT TCAACCGCT ATTGGCTTAA CTGAGCGTGG CATTGTGTTA	240
TCCGCTCCAC AAATTGATAA ATTGCTACAG AAAACTAAAG CAGGCCAAGC ATTAGGTTCT	300
GCCCAAASCA TTGTACAAAA TGCAATAAA GCCAAACTG TATTATCTGG CATTCAATCT	360
ATTTTAGGCT CAGTATTGGC TGAATGGAT TTAGATGAGG CCTTACAGAA TAACAGCAAC	420
CAACATGCTC TTGCTAAAGC TGGCTTGGAG CTAACAAATT CATTAAATTGA AAATATTGCT	480
AATTCAGTAA AAACACTTGA CGAATTTGGT GAGCAAATTA GTCAATTTGG TTCAAAACTA	540
CAAAATATCA AAGGCTTAGG GACTTTAGGA GACAACTCA AAAATATCGG TGGACTTGAT	600
AAAGCTGGCC TTGGTTTAGA TGTTATCTCA GGGCTATTAT CGGGCGCAAC AGCTGCACTT	660
GTACTTGCAAG ATAAAAATGC TTCAACAGCT AAAAAAGTGG GTGCGGGTTT TGAATTGGCA	720
AACCAAGTTG TTGCTAATAT TACCAAAGCC GTTCTCTCTT ACATTTTAGC CCAACGTGTT	780
GCAGCAGGTT TATCTTCAAC TGGGCCTGTG GCTGCTTTAA TTGCTTCTAC TGTTCTCTT	840
GCGATTAGCC CATTAGCATT TGCCGGTATT GCCGATAAAT TTAATCATGC AAAAAGTTTA	900
GAGAGTTATG CCGAACGCTT TAAAAATTA GGCTATGACG GAGATAATTT ATTAGCAGAA	960
TATCAGCGGG GAACAGGGAC TATTGATGCA TCGGTTACTG CAATTAATAC CGCATTGGCC	1020
GCTATTGCTG GTGGTGTGTC TGCTGCTGCA GCCGGCTCGG TTATTGCTTC ACCGATTGCC	1080
TTATTAGTAT CTGGGATTAC CGGTGTAATT TCTACGATTC TGCAATATTC TAAACAAGCA	1140
ATGTTTGAGC ACGTTGCAA TAAATTCAT AACAAATTG TAGAATGGGA AAAAAATAAT	1200
CACGGTAAGA ACTACTTTGA AAATGTTAC GATGCCCGTT ATCTTGCGAA TTTACAAGAT	1260
AATATGAAAT TCTTACTGAA CTTAAACAA GAGTTACAGG CAGAACGTGT CATCGCTATT	1320
ACTCAGCAGC AATGGGATAA CAACATTGGT GATTAGCTG GTATTAGCCG TTTAGGTGAA	1380
AAAGTCCTTA GTGCTAAAGC CTATGTGGAT GCGTTTGAAG AAGGCAAACA CATTAAAGCC	1440
GATAAATTAG TACAGTTGGA TTCCGCAAAAC GGTATTATTG ATGTGAGTAA TTCGGGTAAA	1500
GCGAAAACTC AGCATATCTT ATTCAGAAGC CCATTATTGA CGCCGGGAAC AGAGCATCGT	1560
GACCGCTAC AACAGGTAA ATATGAATAT ATTACCAAGC TCAATATTAA CCGTGTAGAT	1620
AGCTGGAAAA TTACAGATGG TGCAGCAAGT TCTACCTTTG ATTTAACTAA CGTTGTTGAG	1680
CGTATTGGTA TTGAATTAGA CAATGCTGGA AATGTAATA AAACCAAAGA AACAAAAATT	1740
ATTGCCAAAC TTGGTGAAGG TGATGACAAC GTATTTGTTG GTTCTGGTAC GACGGAAATT	1800
GATGGCGGTG AAGGTTACGA CCGAGTTCAC TATAGCCGTG GAAACTATGG TGCTTTAACT	1860
ATTGATGCAA CCAAAGAGAC CGAGCAAGGT AGTTATACCG TAAATCGTTT CGTAGAAACC	1920
GGTAAAGCAC TACACGAAGT GACTTCAACC CATACCGCAT TAGTGGGCAA CCGTGAAGAA	1980
AAAATAGAAT ATCGTCATAG CAATAACCAG CACCATGCCG GTTATTACAC CAAAGATACC	2040

TTGAAAGCTG TTGAAGAAAT TATCGGTACA TCACATAACG ATATCTTTAA AGGTAGTAAG 2100
 TTCAATGATG CCTTTAACGG TGGTGATGGT GTCGATACTA TTGACGGTAA CGACGGCAAT 2160
 GACCGCTTAT TTGGTGGTAA AGGCGATGAT ATTCTCGATG GTGGAAATGG TGATGATTTT 2220
 ATCGATGGCG GTAAAGGCAA CGACCTATTA CACGGTGGCA AGGGCGATGA TATTTTCGTT 2280
 CACCGTAAAG GCGATGGTAA TGATATTATT ACCGATTCTG ACGGCAATGA TAAATTATCA 2340
 TTCTCTGATT CGAACTTAAA AGATTTAACA TTTGAAAAG TTAACATAA TCTTGTCATC 2400
 ACGAATAGCA AAAAAGAGAA AGTGACCATT CAAACTGGT TCCGAGAGGC TGATTTTGCT 2460
 AAAGAAGTGC CTAATTATAA AGCAACTAAA GATGAGAAAA TCGAAGAAAT CATCGGTCAA 2520
 AATGGCGAGC GGATCACCTC AAAGCAAGTT GATGATCTTA TCGCAAAAGG TAACGGCAAA 2580
 ATTACCCAAG ATGAGCTATC AAAAGTTGTT GATAACTATG AATTGCTCAA ACATAGCAAA 2640
 AATGTGACAA ACAGCTTAGA TAAGTTAATC TCATCTGTAA GTGCATTTAC CTCGTCTAAT 2700
 GATTCGAGAA ATGTATTAGT GGCTCCAAC TCAATGTTGG ATCAAAGTTT ATCTTCTCTT 2760
 CAATTTGCTA GGGGATCTCA GCATTGGAGC TACGGCCTGC GCCCTGGCTA AGGATCC 2817

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2861 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2853

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGGCTACTG TTATAGATCT AAGCTTCCCA AAAACTGGGG CAAAAAAAT TATCCTCTAT 60
 ATTCCCCAAA ATTACCAATA TGATACTGAA CAAGGTAATG GTTACAGGA TTTAGTCAAA 120
 GCGGCCGAAG AGTTGGGGAT TGAGGTACAA AGAGAAGAAC GCAATAATAT TGCAACAGCT 180
 CAAACCAGTT TAGGCACGAT TCAAACCGCT ATTGGCTTAA CTGAGCGTGG CATTGTGTTA 240
 TCCGCTCCAC AAATTGATAA ATTGCTACAG AAAACTAAAG CAGGCCAAGC ATTAGGTTCT 300
 GCCGAAAGCA TTGTACAAAA TGCAAAATAA GCCAAACTG TATTATCTGG CATTCAATCT 360
 ATTTTAGGCT CAGTATTGGC TGGAAATGGAT TTAGATGAGG CCTTACAGAA TAACAGCAAC 420
 CAACATGCTC TTGCTAAAGC TGGCTTGGAG CTAACAAATT CATTAATTGA AAATATTGCT 480
 AATTCAGTAA AAACACTTGA CGAATTTGGT GAGCAAATTA GTCAATTTGG TTCAAAATA 540
 CAAAATATCA AAGGCTTAGG GACTTTAGGA GACAAACTCA AAAATATCGG TGGACTTGAT 600

AAAGCTGGCC TTGGTTTAGA TGTTATCTCA GGGCTATTAT CGGGCGCAAC AGCTGCACTT	660
GTACTTGCAG ATAAAAATGC TTCAACAGCT AAAAAAGTGG GTGCGGGTTT TGAATTGGCA	720
AACCAAGTTG TTGGTAATAT TACCAAAGCC GTTTCTTCTT ACATTTTAGC CCAACGTGTT	780
GCAGCAGGTT TATCTTCAAC TGGGCCTGTG GCTGCTTTAA TTGCTTCTAC TGTTTCTCTT	840
GCGATTAGCC CATTAGCATT TGCCGGTATT GCCGATAAAT TTAATCATGC AAAAAGTTTA	900
GAGAGTTATG CCGAACGCTT TAAAAATTA GGCTATGACG GAGATAATTT ATTAGCAGAA	960
TATCAGCGGG GAACAGGGAC TATTGATGCA TCGGTTACTG CAATTAATAC CGCATTGGCC	1020
GCTATTGCTG GTGGTGTGTC TGCTGCTGCA GCCGGCTCGG TTATTGCTTC ACCGATTGCC	1080
TTATTAGTAT CTGGGATTAC CGGTGTAATT TCTACGATTC TGCAATATTC TAAACAAGCA	1140
ATGTTTGAGC ACGTTGCAAA TAAATTCAT AACAAATTG TAGAATGGGA AAAAAATAAT	1200
CACGGTAAGA ACTACTTTGA AAATGGTTAC GATGCCCGTT ATCTTGCGAA TTTACAAGAT	1260
AATATGAAAT TCTTACTGAA CTTAAACAA GAGTTACAGG CAGAACGTGT CATCGCTATT	1320
ACTCAGCAGC AATGGGATAA CAACATTGGT GATTTAGCTG GTATTAGCCG TTTAGGTGAA	1380
AAAGTCCTTA GTGGTAAAGC CTATGTGGAT GCGTTTGAAG AAGGCAACA CATTAAAGCC	1440
GATAAATTAG TACAGTTGGA TTCGGCAAAC GGTATTATTG ATGTGAGTAA TTCGGGTAAA	1500
GCGAAACTC AGCATATCTT ATTCAGAACG CCATTATTGA CGCCGGGAAC AGAGCATCGT	1560
GAACGCGTAC AAACAGGTAA ATATGAATAT ATTACCAAGC TCAATATTAA CCGTGTAGAT	1620
AGCTGCAAAA TTACAGATGG TGCAGCAAGT TCTACCTTTG ATTTAACTAA CGTTGTTTACG	1680
CGTATTGGTA TTGAATTAGA CAATGCTGGA AATGTAATA AAACCAAAGA AACAAAAATT	1740
ATTGCCAAAC TTGGTGAAGG TGATGACAAC GTATTTGTTG GTTCTGGTAC GACGGAAATT	1800
GATGGCGGTG AAGGTTACGA CCGAGTTCAC TATAGCCGTG GAACTATGG TGCTTTAACT	1860
ATTGATGCAA CCAAAGAGAC CGAGCAAGGT AGTTATACCG TAAATCGTTT CGTAGAAACC	1920
GGTAAAGCAC TACACGAAGT GACTTCAACC CATACCGCAT TAGTGGGCAA CCGTGAAGAA	1980
AAAATAGAAT ATCGTCATAG CAATAACCAG CACCATGCCG GTTATTACAC CAAAGATACC	2040
TTGAAAGCTG TTGAAGAAAT TATCGGTACA TCACATAACG ATATCTTTAA AGGTAGTAAG	2100
TTCAATGATG CCTTTAACGG TGGTGATGGT GTCGATACTA TTGACGGTAA CGACGGCAAT	2160
GACCGCTTAT TTGGTGGTAA AGGCGATGAT ATTCTCGATG GTGGAAATGG TGATGATTTT	2220
ATCGATGGCG GTAAAGGCAA CGACCTATTA CACGGTGGCA AGGGCGATGA TATTTTCGTT	2280
CACCGTAAAG GCGATGGTAA TGATATTATT ACCGATTCTG ACGGCAATGA TAAATTATCA	2340
TTCTCTGATT CGAACTTAAA AGATTTAACA TTTGAAAAAG TTAAACATAA TCTTGTATC	2400
ACGAATAGCA AAAAAGAGAA AGTGACCATT CAAACTGGT TCCGAGAGGC TGATTTTGCT	2460

AAAGAAGTGC CTAATTATAA AGCAACTAAA GATGAGAAAA TCGAAGAAAT CATCGGTCAA 2520
AATGGCGAGC GGATCACCTC AAAGCAAGTT GATGATCTTA TCGCAAAAGG TAACGGCAAA 2580
ATTACCCAAG ATGAGCTATC AAAAGTTGTT GATAACTATG AATTGCTCAA ACATAGCAAA 2640
AATGTGACAA ACAGCTTAGA TAAGTTAATC TCATCTGTAA GTGCATTTAC CTCGTCTAAT 2700
GATTTCGAGAA ATGTATTAGT GGCTCCAAC TCAATGTTGG ATCAAAGTTT ATCTTCTCTT 2760
CAATTTGCTA GGGGATCTTG CAACATTGTG CCTGTGAGCA TTGTGAGCCG CAACATTGTG 2820
TACACCCGCG CGCAACCTAA CCAAGACATT GTGTAGGATC C 2861

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "The amino acid at this location can be either Lys, Asp, Val or Asn."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /note= "The amino acid at this location can be either Lys, Asp, Val or Asn."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Gly Xaa Gly Xaa Asp
1 5